A Finite Mixture Approach for Household Residential Choices

by

Michael Shiroya, BBA

A Thesis

In

AGRICULTURAL AND APPLIED ECONOMICS

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

MASTER OF SCIENCE

Dr. Michael Farmer
Chair of Committee

Dr. Eric Belasco

Dr. Emmitt Elam

Dr. Benaissa Chidmi

Ralph Ferguson
Dean of the Graduate School

December, 2010

ACKNOWLEDGMENTS

First and foremost I thank God for bringing me to this point, for His graces, for His presence, for His blessings, without Him I would be nowhere; as my grandmother and mother, two of the greatest mothers in modern history, would say: Nyasaye yetsomwe. I would also like to thank and acknowledge my family for their support, guidance and love, especially my parents. You have sacrificed so much for us and taught us so much; nothing compares to the example you have been to us and the lessons you gave us. We are blessed to have you as our parents. None of us would be where we are without you. I do this and want to be good at this profession to make a difference and so that I can be a reflection of the ideals you instilled in me to the world. Thank you for everything, you are infinitely appreciated. Mad love to Makeba, Kwame and Jimmy, you guys are my mentors and I look up to you guys always. When I grow up I want to be just like you guys; obulala bulera tsingufu - all we have to do is follow the words of mama: tinya! Abeingo stand up, jea!To Cicely, man we've been through so much, good times and bad times, my apologies for my contribution to the bad times and thank you for the good times. You have no idea how much your presence in my life is a source of strength and how you continue to motivate me to strive to be better. I will forever count you as a blessing in my life; I got you for life C. I love you all. Big up to the whole G-squad, we're going to do big things my people, grand hustle!

To Dr. Michael Farmer, thank you for your patience, guidance, support and constantly inspiring me to think outside the box, I hope I end up being half the economist you are. To Dr. Cliff Lipscomb, thank you for believing in me and encouraging me to find my strengths your work ethic, dedication and hard work are unparalleled at the same time inspiring and are a blueprint for success anywhere.

Thank you all.

TABLE OF CONTENTS

ABSTRACT

In the housing sector, on the demand side, attempts to characterize the nature of housing demand have been primarily implemented through hedonics whereby hedonic price functions relate market value to residential housing stock. Hedonic price models, however, do not identity sub-markets and thus may under represent the valuation of amenities and dismaenties in a housing market. Identification of sub-markets allows us to have a more precise and reliable understanding of housing markets in general. The challenge, then, becomes how to identify market segments and impute the marginal value of housing characteristics in different market segments. In this paper, I will attempt to identify and delineate residential sub-markets and derive the predicted marginal values of attributes mixes within those sub-markets. These sub-markets can be thought of composing a homogenous „type' of households.

Sorting households into distinct types using a statistically robust approach is a useful way of accounting for preference differences in a structurally sound manner that offers deeper insight into the consumption and valuation of different amenity packages. In this study, I successfully identify and differentiate types of households (sub-markets) by implementing a finite mixture model. The extensions of the concepts and analytical findings of this study can be applied to model heterogeneity in larger studies with moderate effort.

LIST OF TABLES

PREFACE

"Yatikhanga niyakhola hamuliango."

Translation,

"Don't be overly sure."

-Dr. O.J.E. Shiroya

CHAPTER I

INTRODUCTION

A key to implementing successful urban redevelopment is to understand the role urban amenities and disamenities play in urban welfare and economic activity. Policymakers and citizens require some efficient methods to value amenities and potential disamenities to compliment certain public ends. Business attraction, development and growth, and the socio-economic make-up of communities have shaped public policy in a manner that crosses geographic scales and political jurisdictions. Therefore, there is efficacy in having a better understanding of the socio-economic typography of communities and how these differences affect the demand for attributes of housing.

This work plans to have a better understanding of a housing market by implementing a conceptual approach that fuses classical and neo-classical theories - on household residential choice patterns, housing attribute valuation and land use patterns - with modern econometric techniques. The motivation in this case is to adapt Hedonic Pricing methods to reflect certain layers of complexity in what makes up an urban area; its economic activity and the role of policy, largely zoning policy and human capital recruitment in this case, in informing how hedonic price analysis can be conducted.

Urban communities are a complex array of decisions: policy related decisions; business related decisions; and residential related decisions. The urban space itself: the housing stock, the commercial and office building stock, public areas and a complex web of infrastructure is forged out of community dynamics shaped in varying degrees

by socio-economic effects. This work focuses on one key complexity in the residential housing market which is arguably a reflection of the broader urban policy, social and economic dynamics – household diversity. In general the thesis argues that fully separate types of persons with fundamentally different preference orderings can co-exist in the same neighborhood. How citizens of different types disburse and allocate urban space is less regimented than is often modeled. Neighborhoods are often not homogeneous as, say, the Tiebout Hypothesis contends; nor do multiple members of a neighborhood order their preference for the next best and next worse neighborhood in the same way. Those nuances are interesting in themselves, but the implications for amenity provision or for economic development or for regional economic planning generally are not trivial.

Social planners, for example, when taking into account the effect of implementing a particular land use change, for example, re-zoning an area for the addition of a school or for the development of a transportation highway; generally assume that the effects or response of that land use change would be the same across all households and proceed to make decision with an expectation of relatively homogenous effects. Now if the assumptions of neighborhood homogeneity are informing into urban land use decisions and provide guidance for targeted urban redevelopment initiatives, then it is prudent from a planning perspective to test if those assumptions hold. Additionally, local governments estimate their funding potential given the socio-economic make-up of their communities, their neighborhoods. As such, if the local amenities preferred by households influence the local tax base, then it becomes an economic priority to have better insight into the reference and valuation

patterns of the households in their jurisdictions. Typically, housing market models are built on under Tieboutian assumptions that would offer local governments and planners a very general idea and in the process they mask useful information that could change the implications of that model; however if economic priorities call for a more refined insight into the working of a housing market then a model that relaxes hedonic model assumptions would bear more useful inferences. The effect of masking important inferential information becomes more pronounced as we collapse the scale of geographic references from the community to the neighborhood and from the neighborhood to the household. This is because urban redevelopment activities , for example, would fail to have the intended impact if the deeper understanding of the workings of a housing market. Overall, failures in public policy initiatives could be avoided if we have a better idea on how households make residential decisions and how they value local amenity mixes, i.e the local residential stock. In the array of decisions, residential decisions, their interaction with available stock in the community, and the outcomes of these decisions play an important role in shaping the contours of a community. In other words, the varying levels of demand for different attributes of housing dictate the distribution of households across a community.

Attempts to characterize the nature of housing demand have been primarily implemented through hedonics whereby hedonic price functions relate market value to residential housing attributes. Residential house value hedonic functions quantify house values relative to its physical and location attributes. This idea is rooted in applied economic theory, particularly consumer utility theories. Samuelson (1954) ideas on consumer utility were based on the assumption of two categories of goods:

private consumption goods and collective consumption; by using choosing this categorization of goods in his approach to consumer utility, his analysis was fittingly conceived from a demand side viewpoint, through a lens of individual preference revelation and consumption. Collective consumption goods were considered to be good which consumers „..all enjoy in common in the sense that each individual's consumption of such a good leads to no sub-traction from any other individual's consumption of that good'. Each individual is expected to have a set of preferences for collective consumption goods, or public goods, as well as for private goods.

Samuelson argued that consumers directly reveal their preference when they consume private goods only; because they can weigh the marginal benefits of an additional unit of a good or an additional attribute unit relative to price parameters. Tiebout (1956) theorized that consumers could indeed reveal their preferences for public and private goods - for the individual attributes of public and private goods by implicitly decompose the good into its separate services.

He stated that, assuming mobility, consumers would reveal their preferences for public goods by „voting with their feet'. Where Samuelson stated that optimal pricing systems could not be explicitly determined for public goods, Tiebout posited that that pricing system is implicitly determined by an unobserved process in which consumers identify, consider and choose different combinations of private and public goods, or bundle services into a single purchase. Tiebout suggested that such unobserved transactions unintentionally mimic market mechanisms as supply differences of public goods across localities result in efficient allocations of resources and parity in household sorting patterns. This diverges from Samuelson's ideas that a

lack of a conventional market, as is in the case with private goods, results in poor inefficient allocation and use of resources.

At the extreme, Tiebout's hypothesis would result in homogenous households coexisting in the same neighborhood. These households would have identical preferences and as such, have identical demand functions. In this case, all the demand functions of these households can be reliably aggregated into a single market demand curve to estimate implicit prices of a neighborhood's amenities. The resultant implicit prices can be considered to be reliable and fully revealing of true value.

But Samuelson's 1954 study involved an important case. He noted that urban spaces were inefficient in the allocation of space because the services or activities in an urban and the then emerging suburban space could never realize the needed scale to provide these services at the „bottom of the average cost curve.' Space constraints, he argued, pre-empted this form of activity efficiency. Yet the tradition of urban settlement patterns that follows Von Thunnen (1826) offers another version of efficiency: the highest and best use of space. Stated another way, one way to judge the efficiency of a space is whether the overall value of space, collectively, realizes a higher value. In this way parks that lift up values of surrounding property can be efficient if those improvements 'outbid' the next best alternative economic use of that space. There is no deep reason to expect the size of the park and its activities would expand to realize efficient scale economies we might expect of commodities sold in a perfectly competitive market. The size and scale of park activities would stop at the marginal value benefit of those activities to all costs, direct provisions costs and foregone opportunity costs for that space and those resources.

However, accounting for household heterogeneity at the neighborhood level unmasks differences in implicit marginal value for neighborhood amenities if a neighborhood is less homogenous than an extreme case of the Tiebout model would dictate. One gross, time-consuming approach toward valuation of amenities in neighborhoods comprised of heterogeneous households would be to analyze demand functions for each individual. This of course echoes Rosen's argument on the hedonic specification problem. Potentially each unit sale is an interaction of a unique supplier and a unique purchaser. So heterogeneity has the potential to break down into almost perfect "one house – one type" total market segmentation.

Some researchers suggest that a satisfactory way of dealing with heterogeneity is by taking these demand functions and accounting for heterogeneity with a shift in intercept in the array of demand functions. However, we should note that from a statistical efficiency point of view, heterogeneity in a data set may not be entirely captured by a shift in intercept for there may be other reasons, unobserved to the researcher, which may be correlated to consumer behavior and choice patterns; statistically this implies that a shift intercept does not capture the possibility that the error terms may be correlated between individuals.

Applying a choice model that allows for intercept heterogeneity only, effectually imposes an implicit restriction on the structure of heterogeneity. In other words, there is still a possibility that the reasons for consumer behavior may not have been fully captured. There is still room for explaining more effectively how consumers are reacting to exogenous factors. The marginal valuations represented by each beta, of a choice model of this form, could still be different for each person; a better

estimation of this would allow us to relax the assumption made that amalgamate consumer choice patterns so as to delineate dissimilar consumer choice patterns and group similar choice patterns. Hedonic models as typically practiced fall short in this regard. Hedonic models were popularized by Rosen (1974) who basically postulated that demand analyses of these types of goods can be derived as a function of that product's characteristics.

For example, a hedonic model applied to the demand in the automobile industry would specify the demand for all cars as a function of, for example, horse power, mileage, and EPA fuel consumption estimates. The hedonic model would use ordinary least squares methodology to regress the price of cars against the aforementioned explanatory variables in a first stage equation with the assumption that this is the correct functional form; thereafter, a second stage equation would compute the partial derivatives of a particular explanatory variable, say horse power, and regress it against other variables that relate to it; this second stage was meant to derive the demand for whichever attribute of interest.

This second stage equation, as Rosen (1974) articulated it, was to be regressed against household *demographic characteristics* so as to capture household preferences, better matching types of persons to different marginal valuations of hedonic amenities. Epple (1987) warns that this process offers inaccurate estimates „regardless of the sources of error in the equations'. Furthermore, Rosen started off developing his theory for this second stage equation by postulating that every single sale was a separate market, so the market differentiates into as many markets as there are sales because each package is unique, each buyer is unique, each seller or supplier

unique, then every observation is its own market and in a sense you cannot aggregate at all. This hypothesis later directly competes with the assumptions of the model he later forwarded to describe residential choices. In his later model Rosen's specification relies on the assumptions of homogeneity, perfect competition and parity in information costs, income and mobility, and a continuous residential stock choice set. Here, a continuous residential stock choice set implies that every household can consume a particular amenity infinitely; invariably implying the existence of a continuous, dense supply of residential stock. If any of these assumptions are violated then any analytic conclusions from Rosen's estimation strategy are inconsistent.

Additionally, this approach does not allow for a flexible implicit price estimates that allow for heterogeneous consumer tastes and preferences to interact with attributes of residential stock in a way that reveals implicit prices at consumer utility maximization levels. Rosen's specification is more appropriate to characterize general trends over large areas. This aggregation problem arises because households' unique preferences for a unique bundle of good cannot be measured uniquely as such Rosen's specification only allow for aggregated demand functions. Hence the outcomes of his specification would be appropriate if *all* households choose to live in neighborhoods for *all* the same reasons. Such a process would follow a Tiebout model of household residential location decisions. Tiebout's model notably offered a partial resolution to the one house-one type implication and the extreme homogeneity implication.

Tiebout's 'voting with your feet' idea had dictated that all the households move into and occupy a neighborhood for all same reasons. Moreover, a central

implication of Tieboutian homogeneity outcomes is homogeneity characterized and manifested by all households in a particular neighborhood ranking all houses in that neighborhood the same. By extension, therefore, if one was to randomly pick a household in a particular neighborhood and ask them to order all houses in the neighborhood from most valuable to least valuable and ask them the reasons for their particular ranking; all other *N* households in that neighborhood will have the same the ordering, the same reasons for that ordering thus the same amenity preference and consumption patterns. However, if the strict Tiebout model is violated, then distribution of household types may not be singular such that there are notable differences amongst households, there is ground to challenge the absolute existence of only one type in a single neighborhood, that is a violation signals the presence of sub markets within a single neighborhood.

Therefore, in this study, I hypothesize that household preferences differ and that they make residential location decisions for different reasons, even in the same neighborhood. I put forward that there could be some dimensions of preferences for specific characteristic bundles of a house that are similar enough among a sub-group so that a finite number of sub groups can be identified such that each can be generalized into a distinct class of household - or household type. I hypothesize that these distinctions are unlikely to obey a continuous distribution among types but rather the distinct classes will be a finite number of quite distinct types. That makes the estimation and sorting process harder econometrically, however it leads to more a robust characterization of household residential sorting patterns.

Given the aforementioned, the objectives of this study are to:

1) Endogenously group households into distinct types.

2) Identify neighborhood diversity profiles that value and support neighborhood amenity packages.

3) Estimate the valuation, by type, of amenity-mix supply packages.

If we assume that households have identical preferences we imply they have identical demand functions. In this case, all the demand functions of these households can be reliably aggregated and a single market demand curve can be used to estimate implicit prices of a neighborhood's amenities; the resultant implicit prices can be considered to be reliable and fully revealing of true value. However, accounting for heterogeneity unmasks differences in implicit marginal value for neighborhood amenities if a neighborhood is less homogenous than an extreme case of the Tiebout model would dictate. Valuation of amenities in neighborhoods comprised of heterogeneous households would require a more surgical analysis of housing demand.

CHAPTER II

LITERATURE REVIEW

2.1 Sub-markets & Segmentation Strategies

One of the classic underpinnings of consumer behavior theory is that consumer decisions are driven by the desire to maximize utility. This assumption forms the foundation of modeling economic scenarios involving consumer decisions. The idea of maximizing utility implies that consumers will choose the combination of goods and services that maximizes their utility. The theory implicitly relies on the assumption that consumers inherently derive different levels of utility from various combinations of goods and services. Building on earlier works from earlier authors such as Quarks (1956) and Stigler (1945), Lancaster (1996) postulated that the level of utility a consumer derives from a good is dependent on the characteristics/properties of that good. Furthermore, due to differences in tastes, preferences and needs amongst consumers, consumers will derive different utils from the same good; as such, consumers place different value levels on combinations of goods and services. In other words, value placed and satisfaction accrued will vary across consumers exposed to the same product bundles. These considerations have been incorporated in much of the literature on consumer behavior since Lancaster and their tenets have also been applied in the analysis of housing markets through hedonic price modeling.

Authors have defined the properties of houses in a myriad of ways, most commonly used characteristics of houses are square footage, lot size, age, bedrooms, bathrooms, garage, swimming pool, fireplace and air conditioning (Sirmans et al., 2006); also the attributes of the geographic area – that housing market - within which

that house is located in i.e external or neighborhood characteristics of the house are included in pricing models (Hughes and Sirmans, 1992). Together, the aforementioned form a housing bundle stock that we cannot separate a house from the land it is built on or the environmental attributes of the neighborhood within which it is located; bearing this in mind, houses then are big pre-bundlings of attributes which are consumed as a component. Therefore, housing models must – borrowing from Lancaster's theory – take into consideration the varying levels of utility derived by different households in a housing market.

Bearing in mind consumer behavior theory, consumers will make residential housing decisions according to the combination of housing characteristics that maximizes their utility. As such, consumers will gravitate toward their utility maximizing locations, forming pockets of households – relatively similar in terms of tastes, preferences and needs as predicted in the Teibout hypothesis. These pockets form within the larger market that would have a different willingness to pay for their utility maximizing combinations than a dissimilar pocket of households within the same housing market. Hedonic models do not capture and valuate these pockets or sub-markets in a robust way because hedonic price models treat an array of house prices as a single unified market (Bitter et al, 2006); resulting in models that approximate effects on average. This approach poses a validity threat as pointed out by Can who asserts that amenity variations across housing markets should not be treated as determinants of housing values in fixed coefficient specifications as in housing econometric valuation models (1990). As such, these empirical models, have failed to take into account the dynamics of sub markets in the analysis of property

values; thus analyses arising from these models include some aggregation error (Riddel, 2001). Palmquist, for one, has contended for some time that this oversight poses a threat to the accuracy of predicted outcomes (Palmquist, 1984; 2004). Generally, the assumptions of unitary housing market functioning (Maclennean and Tu,1996), are violated if there are submarkets within the respective housing market under study.

Bourassa et. al (2003) and Goodman and Thibodeau (2003, 2007) emphasize that the relevance of market segmentation lies with the premise that the accuracy of price functions is improved when sub market dynamics are taken into consideration. Though the idea of housing sub markets has been talked about for decades, for example since Grisby (1963), and Straszheim (1975), it was mostly isolated to the fringes of housing research and did not gain popularity in housing research until the more recent past. However, despite the growing body of literature in this domain of housing research, a common definition of what a sub market has yet to be penned. Researchers have derived sub markets using differing criteria and estimated effects in sub markets using different econometrics approaches. Bourassa et al, (2003) note that the methodological approach to segmentation depends on the impetus of the researcher; for example, studies that are solely geared toward understanding price distributions or predicting housing stock valuations differ in their identification methods. Notably, however, much of the literature uses some form of the hedonic approach in the process of their segmentation analysis.

Newsome and Zietz (1992), suggest the existence of variations in a given distribution of housing prices.In this vein, existing literature reflects that housing sub

markets are assumed to arise due to different factors- either price-related effects, or spatial/geographic factors, or housing attribute characteristics, household characteristics/ preferences or some combination of the factors. The following sections will outline the different approaches to segmentation – how researchers have classified sub markets and how they have estimated effects in those submarkets.

Costello (2001) and Rothenberg (1991) approach the segmentation process by relying heavily on price-related changes in the data set. This approach effectually assumed that sub markets formation is purely price driven. As such, this approach would derive a hierarchy of prices and rank households into sub markets based on price bounds or attributes of housing stock within those bounds. This approach has parallels to truncating data based on a condition and separately evaluating the abridged data set. Heckman (1979) highlights the errors of truncating and using price boundaries as the sole segmentation criterion (Newsome and Zietz, 1992) in housing valuation models.

Other researchers have defined submarkets using benchmarks such as proximity to schools, employment, parks, transportation or central business districts/city centers. These approaches delineate sub markets based on the distribution of prices changes from proximity to the benchmark of interest. These approaches solely view market structure as a pattern of price changes radiating from a focal point. The econometric extension of these approaches would be to place each household in a group based on a distance criteria. Fik et. al(2009), Abraham et al(1994), and Kauko and Goetgoluk (2005) incorporated proximity/distance gradients in their identification of submarkets. Literature that approaches segmentation in this manner typically

derive valuation estimates that are not constant over space, as such, the markets under study are concluded to be geographical divided. However, this approach assumes homogeneity at the household level within those geographically-derived groups. A closely followed approach, defines sub markets according to the quality of dwelling units only or as a nested function of geographic segmentation and dwelling unit qualities segmentation. Researchers such as Sirmans et al (2006), Leishman (2001) and Watkins (2001) shadow this approach in their analyses of segmentation.

Zhuo et al. (2009) explain that property prices in sub-markets are determined by different functional relationships; functional relationships that involve the interaction of housing demand and supply factors that affect a consumer's willingness to pay. The functional relationships between the structural component of housing, amenity externalities, and household preferences make up the structure of a housing market. These interactions are such that sub markets are not only segmented due to price-related factor or dwelling unit quality & geographic related factors solely but as a combination of multiple factors that affect the whole package of housing services.

In this regard, Tu (1997), Goodman and Thibodeau (2007) propose a multi-level approach where segmentation occurs at two levels – geographically and then based on unit quality at the intra geographical group level. The majority of the literature that is in line with this approach implement multiple econometric procedures to derive the hierarchical segmentations. Bourassa et al (1999) use principal components, k-means clustering and OLS in their segmentation approach. They use principal components to deduce, from a hypothesized set of factors, the factors would be the most effective in delineating market structure, and then used these factors to the

number of sub markets that existed in the relevant area of study through k-means clustering, and then run regression in each sub market for valuation comparisons. This step-wise approach is mimicked by other researchers who use different econometric approaches to successively derive sub markets. For example, Dale-Johnson (1982) use Q-factor analysis & OLS whereas Adair et al (1996) use OLS and the F test.

The aforementioned approaches fall short in capturing and accounting for household preference heterogeneity. To truly capture household heterogeneity, one has to employ a discrete choice approach that seeks to identify the marginal valuation of a product's attributes to a purchasing decision. Such a methodological presentation would not only provide useful information on the nature of residential choice but also reveal relations between housing stock and occupancy of housing stock that holds at all levels of aggregation.

Much of hedonic price analysis which searches for the value of housing amenities, treat very large groups as a single market. If individual demand functions are much more different (i.e. not agreeable), then it becomes an economic development concern to examine the reasons for the differences. A consumer's tastes and preferences shape their purchasing decisions, which are represented by an individual demand function. Consequently, if we dissect individual demand functions within a particular market to better understand the desires and responsiveness of different consumers in that market, then we can better understand the overall market. Expressed differently, it is important to identify the varying levels of importance consumers place on various attributes of commodities or commodity bundles.

2.2 Heterogeneity & Discrete Choice Models

Since demand is a function of preference heterogeneity, it is important to include preference heterogeneity in the function relationships that inseparably contribute to valuation of residential housing stock. Indeed some scholars have implemented approaches to deal with individual heterogeneity in the housing purchasing decision. Discrete choice models implemented in redistill markets are built on theory of McFadden's logit model family; such as those from Quigley (1976) to more recent work from De-Palma (2007) and Bruch and Mare (2009). The application of the family of logit models is useful in studying the effects of individual heterogeneity in housing purchasing decisions. Discrete models, in essence, seek to predict the probability of making a certain choice. They all assume that a consumer, faced with a set of alternatives, will choose the alternative that gives them the most utility. Discrete models, in essence, seek to predict the probability of making a certain choice. They all assume that a consumer, faced with a set of alternatives, will choose the alternative that gives them the most utility.

They are many forms of discrete choice models. Researchers have used discrete choice models to estimate the distribution of demand and willingness to pay for attributes of various goods and services (Teratanavat & Hooker, 2006 and Kwak et al, 2010). The most commonly implemented models, as aforementioned, being logit and probit models and their related derivatives. Though there many forms of logit models, they generally fall into three categories: models whose coefficients are the same across all choices (multonomial logit models) those whose coefficients are a function of the characteristics of a specific choice (conditional logit models) and those

whose coefficients are of both kinds (mixed models). However, these models are more informative about the probability of choosing a particular house; however they are not as informative of the structure of unobserved heterogeneity and how it impacts valuations of attributes of residential stock. For example, a conditional logit model would not allow for variation based on demographic factors but rather assume that it is fixed at the same value for each consumer as in. Scholars have attempted to capture these affects in various ways, for example, Lipscomb and Farmer (2005), use a seemingly unrelated regression process to account for unobserved individual preference heterogeneity followed by iterative hedonic equations to identify submarkets. A more formal computational approach would be to implement a form of discrete choice models that simultaneously estimates these effects.

In this paper, I attempt to implement a model that builds on the foundations of Lipscomb and Farmer (2005) in that I seek to model how individual heterogeneity interacts with attributes of housing stock at resulting in observed housing sale transactions. In this regard, I employ a form of latent class analysis. Latent class analysis (Lazarsfeld and Henry 1968) can be used to identify groups or clusters that best account for patterns of vocational decisions in the array of household data. Latent class analysis has been used in wide array of fields including health economics (Thatcher et. al, 2005), microbiological sciences (Nielson et al, 2004), and marine sciences (Wallmo and Edwards, 2008).

Latent class models are more appropriate for this study because they rely on a mixing distribution of observed features that likely signal the presence of different types of choice conditions. Latent class models are capable of managing multiple

simultaneous conditioned effects on parameter estimates, large diversity within a particular study group and are capable of managing relatively small data points by extrapolating by simulations from a few surveys to capture variation.

CHAPTER III

CONCEPTUAL FRAMEWORK

3.1 Household Residential Choice Patterns

Recall, Tiebout asserted that marginal utilities are practically revealed by the consumer-vote, where their voting behavior is viewed as a disclosure of their optimal consumption bundle vis-a-vis their preferences and taste. Tieboutian inspired hidden market processes claim demand revealing processes and resulting endogenous market clearing prices.

The idea of hidden market mechanisms driving observable transactions favors the concept of a Vickery style auctions. Vickery auction mechanisms have been used to form the theoretical rationale of research attempts to valuate consumer's willingness to pay for varying food attributes (Alfnes, 2007a), examine behavior in job assignment (Nimon et. al, 2007) and contraction contract auctions (Drew et. al 2006). Vickery auction processes where posited by William Vickery (1961) as an auction process by which a good is sold to the highest bidder at the second highest bid price. Vickery auction bid prices are theorized to reflect people's true bid price for consumers because the auction process matches them to what they have to actually pay; as such over-bidders risk utility losses, while under-bidders risk foregone utility(Alfnes,2007b). Since consumers are expected to act in a way that maximizes utility in economic transactions, consumers reveal their private preferences by offering true bid price or willingness to pay, a price that maximize their utility given their private preferences. If consumers face diminishing marginal values characterized by monotonic demand functions then the optimal clearing price for a commodity bundle

for each individual is the point at which the gradient of a consumer's utility function equals the market price. At this point, under/over-bidding would be non-existent and the consumer would have maximized their utility relative to their preference and budget constraints.

Note that Tiebout formulated his hypotheses based on a set of assumptions including: full mobility of consumers, a large number of communities for communities to choose from, and that there exists an optimal size of a community (Tiebout, 1956). Given the aforementioned assumptions, consumers would efficient self-allocate to the neighborhoods that best satisfy their preferences. Therefore, assuming two consumers *C* and *M* are weighing neighborhood alternatives and the aforementioned Tiebout assumptions apply for the two consumers, i.e. full mobility and a large number of communities to choose from. Additionally, the Tiebout model relies on the idea that households are able to choose from a large number of neighborhoods so that if a household's preferences are not prevalent in that neighborhood they simply move to another fitting neighborhood. However this manner of sorting as a result of an implied abundance in household production can only be practically achieved assuming an unlimited amount of land (Tideman, 1998).

Standard models of land allocation, dating back to Johann Heinrich von Thünen say that space itself is scarce. Von Thünen's regional land use model is on of the oldest regional land use models that highlight the effects of land scarcity. He developed a model to explain the pattern of land usage of different agricultural activities; with the core assumption that land use follows a pattern of concentric circles around a market. The principles of this model dictate that since space itself is a

constraint then potential production activities compete for space in the most valued land, which in this case was land that was closest to a market. The underlying implication of this model are that given space constraints influence land use patterns.

The pure Tieboutian consumer-vote model would, as Tiebout premised, lead to consumers with the same preferences ordering for residential attribute mixes all settling in the same neighborhood. This outcome in his model relies on the existence of a large number of communities for households to choose from hence the greater possibility and feasibility for a single type of household to comfortably settle into a single neighborhood. This outcome of homogeneity may not always be reached given limitations in land supply. Given land supply limitations, what is absent in that purely Tieboutian driven models is the consideration that the space required to always have one neighborhood-one type may not always exist; for the space required for just one type to comfortably occupy a neighborhood is space that other types could want as well since other types are also forced to seek that same space due to limited supply of parcels. Types of household compete for the same space. The space constraint, especially in downtown neighborhoods where the effects of a space constraint are experienced more, does not allow for a infinite number of neighborhood choices. Hence, the formation of a uniform distribution of a type of household in a single neighborhood of optimal size may not always be possible. Incorporating implications of von Thünen theory implies that a homogenous neighborhood is not always guaranteed. Some neighborhoods could be characterized by a single type, some may not.

Rosen touches on an extreme outcome of space constraints. He started off his hypothesis by implying that the number types in a neighborhood would be the N number of households in that neighborhood; and then retracts from this conjecture by forwarding that since there is no way to deal with that extreme level of heterogeneity we should assume a one-neighborhood, one-type model of household sorting. The quandary of modeling household residential sorting patterns shifts from extreme heterogeneity to homogeneity. However, neither two extremes are appropriate even when one-neighborhood, one-type allows for some diversity.

Johann Heinrich von Thünen's model of land allocation articulated land allocation patterns as concentric circles radiating from a market that essentially was a focal point. Here, the land use activities that were closest to the market had the highest marginal productivity and paid a premium for being close to the market; the land use activity that had the highest economic rent outbid other activities for a more favorable location. Similarly, types of households compete for a limited supply of housing amenities, or rather a limited attribute mix supply; hence households try to outbid each other for particular parcels. This process, I hypothesize, partly leads to the violation of a consistent outcome of homogenous neighborhoods and instead allows for different types of households to occupy the same neighborhood and that there are inter-type differences in the ranking of neighborhoods and houses in a neighborhood – from the most valuable (or desired) to the least valuable; the implications of the aforementioned makes a difference in how one would understand and model amenity behavior.

Note that in the process of a house purchase transaction, each transaction involves a bidding process. Vickery auction theory, when applied to housing

transactions, dictates that when one is bidding for a house, what really matters is that you outbid the second highest bidder by some epsilon *x*. The price does not always reflect the true willingness to pay for the highest bidder, it is a sharper reflection of true willingness to pay for the second highest bidder. I hypothesize that, since we are only observing the second highest bid, it is likely that there are competitive bidders in a market who are relatively similar; that is, there is a group of people who are competitively bidding at or near that second bid who are relatively small in number (compared to the entire market) and that they probably have similar preferences - by revelation - because they are willing to pay similar amounts for similar packages. Out of all the households, it is that group of people who are placing competitive bids for a particular attribute mix hence there is a level of similarity within that group. I assume that the very unique preferences of a particular competitively bidding household is barely visible in the price array as such I forward that if there is a lot of orderliness in a market, it is likely that the preferences of competitive bidders are relatively agreeable; and can be collectively characterized with reasonable precision and can be statistically clustered into a type.

Bearing this in mind, in this study I allow for a housing market to sort based on the differences in the characteristics and preferences of individual households for attribute mixes and investigate how well those differences translates into different types. The process of sorting is determined by the characteristics of the goods themselves and the preferences that consumers have for those goods. My approach allows for variety in the allocation of sub-markets across space, it doesn’t force sub-market formation into geographic or income level bounds.

Different types living in the same neighborhood implies that people with different preferences are consuming similar amenity packages but differently, which means that these types could have different orderings in their hedonic indexes. So for example, two house prices could be the same, but since they represent different houses, say houses *H* and *W* respectively, they could represent different amenity packages; hence there could be one group of people bidding for the amenity package of *H* while another bidding for the amenity package of *W* – thus, for example, segmentation on price alone wouldn't be as efficient and informative. There would be households that value the particular amenity package offered by *H*, while others that value the amenity package offered by *W* such that they value and order the amenity package differently and substitute between the components of the amenity package differently.

The trajectory of valuation of components of the amenity packages between the two households for *H* and *W* would be different. Thus the households display different hedonic indexes. So, if we assume that the two households represented two different groups - groups *V* and *G* - and that *H* and *W* represented the most valued parcels by *V* and *G* respectively, note that the highest bidders in *V* and *G* would acquire *H* and *W*. Accordingly, however, the trajectory of square footage, for example, should have a consistent improvement between the poor and the wealthy in *V;* which could look different from the trajectory of square footage from the poorest to wealthiest in *G*. Finite mixture approaches are the econometric extension of trying to tease out such differences among households in a statistically sound manner in order to group them into a type.

The output of integrating the aforementioned theoretical viewpoints with discrete choice analysis allows for the possibility of homogenous neighborhoods but also acknowledging the likelihood for mixed neighborhoods; that is, it allows for multiple types to exist both at the inter- and intra-neighborhood levels, essentially challenging traditional approaches to modeling residential housing choices.

3.2 Latent Class Analysis: The Finite Mixture Approach

The traditional functional form for regression models assigns a weighted coefficient to a given characteristic as that characteristics' marginal contribution to the price of the house; the functional form of this being:

$$\text{Price} = \beta_0 + \beta_1 X_1 + \ldots. + \beta_N X_N \qquad (3.2.1)$$

The above equation is interpreted as being representative of the relative value (β coefficients) of the characteristics of the house – i.e housing stock attributes - which are collectively capitalized into price of the house. The price of the house as a gauge of consumer willingness to pay is captured by the interaction of the housing stock attributes as defined by their fixed parameters. The parameters are not robust because they do not allow for effects that are individual variant; thus one has to consider how to model the differences in individual choice behavior that is not captured by X_N.

This consideration would not matter if consumers have the exact same preferences. However, it matters if consumers have different preferences and have different rankings for their preferences; that is, if they assign different relative values to the various attributes of the house which in turn dictates their willingness to pay for the house. In essence, the β coefficients in the above equation are an average of the fixed parameters of all individuals in a particular study. At best, they represent a

weighted average of consumer responsiveness to various housing attributes. This aggregation is less efficient when heterogeneity exists among consumers. Some researchers suggest that a satisfactory way of dealing with heterogeneity is by taking household demand functions and accounting for heterogeneity with a shift in intercept in the array of demand functions.

However, we should note that from a statistical efficiency point of view, heterogeneity in a data set may not be entirely captured by a shift in intercept for there may be other reasons, unobserved to the researcher, which may be correlated to consumer behavior and choice patterns; statistically this implies that a shift intercept does not capture the possibility that the error terms may be correlated between individuals.

Applying a choice model that allows for intercept heterogeneity only effectually imposes an implicit restriction on the structure of heterogeneity. In other words, there is still a possibility that the reasons for consumer behavior may not have been fully captured. There is still room for explaining more effectively how consumers are reacting in to exogenous factors; a better estimation of this would allow us to delineate dissimilar consumer choice patterns. An approach that gives us better information on the structure of heterogeneity and more reliable understanding of the consumer responsiveness to various combinations of attributes would be more appropriate. To statistically accomplish this, one should employ a form of discrete choice models.

One possible discrete choice approach would be to to adopt a random parameters model. The growing practice of a random parameters model (Tu &

Goldfinch, 1996, and Suzumura & Xu, 2003) does offer some hope here; yet there is a strong assumption of a continuous distribution of diverse types that in many cases is violated (Farmer, Belasco, and Lipscomb, 2009). A random coefficients model is an extension of mixed models. It results in coefficients that vary within the population – and in this paper, variation is based on demographic factors - rather than fixed at the same value for each consumer as in , for example, conditional logit models.

The general framework of the methodology of random coefficients models first involves the specification of an underlying distribution of the attribute parameters to be estimated. From this entire distribution of attribute parameters, a distribution of estimated attribute beta coefficients, based on likelihood, for individuals is generated. In theory, this distribution is of an infinite number of points; by extension, this implies that every household would be its own type with is its own set of parameter estimates since each point on the distribution could be taken to represent a different household.

However, we forward that it is also possible for a set of parameter estimates to be close enough - as a certain group of individuals having similar enough preference rankings - which means that instead of having an infinite number of points to represent each household or individual one could observe a grouping of points such that a density is formed along certain sections of the distribution. These densities would be considered a grouping of similar enough individuals to collectively form a type or class of household; such a scenario would parallel the theoretical outcomes of the latent class analysis of a finite mixture model. In this way the random parameters model is a special case of heterogeneity that obeys a strong assumption of a continuous distribution of heterogeneity across a target population.

Latent class models rely on a mixing distribution of observed features that likely signal the presence of different types of choice conditions. These models manage multiple simultaneous conditioned effects on parameter estimates, large diversity within a particular study group and are capable of managing relatively small data points by extrapolating by likelihood from a few surveys to capture variation. The structure of variation is used to classify data points into a class.

Latent class analysis finds classes of observations that are similar based on observed characteristics, identifies the characteristics that delineate groupings, and estimates the prevalence of classes. Finite mixture models provide ways to account for unobserved heterogeneity (Kashara et al, 2009). The finite mixture approach seeks to account for preference and structural heterogeneity endogenously. As such, the individual preferences and choice response structure are characterized for each cluster. Since finite mixture approaches decompose an array of choices into component choice sets, with each component choice set describing a particular class, they implicitly assume that the mixture component functions have a particular distribution. The parameters of these distributions and the density of each class are estimated to reveal class density functions. Dempster et. al (1977) first proposed the estimation of these parameters using an Expectation-Maximization (EM) Algorithm. This estimation technique is implemented to test for and account for unobserved heterogeneity in the intercept and explanatory variables. The EM algorithm finds a maximum likelihood estimate of the parameters of the class probabilities that produced observed outcomes; the algorithm accomplishes by randomly classifying observations into classes, and successively regrouping observations into temporary groups based on a specified

improvement criterion until the criterion has been satisfied. Friedman (1998), followed by other authors such as Elidan et al(2001) have forwarded the EM algorithm as a way to reveal the structural effect of latent variables in a data set. Their impetus, and that of other researcher who have used the EM algorithm in their methodological approaches, is driven by the view that incorporating the effect of latent variables is key to improving the understanding of interactions within a data set which sheds a light on hidden market mechanisms.

Recall that at the extreme, Tiebout's hypothesis would result in homogenous households coexisting in the same neighborhood. Bearing the aforementioned in mind, embedded in a methodological approach that relaxes some of Tiebout assumption is an inherent test of heterogeneity. This paper approaches this housing valuation study from the viewpoint that given a relaxing of Tiebout assumptions, heterogeneous household sorting could exist. Given that we expect a heterogeneous residential housing market, by extension, we expect a mixed distribution of household preferences. As we hypothesize a mixed distribution of households, we expect households to respond differently to observable differences in housing stock.

Our aim is to examine a distribution of residential housing stock demand to identify different types of households and the characteristics that define them. Latent class models assume discrete distribution of unobserved heterogeneity in their analysis of sample data. They assume the presence of stratified homogenous sub-populations within a larger population; such that a sample drawn from the true population contains a finite number of distinct sub-populations which are considered to be independent of each other. Therefore, each population is theorized to be made up of *C* sub-

populations and likewise, each a sample of the population is also assumed to contain *C* number of classes – a representative of the larger sample.

CHAPTER IV

METHODS

The stratification of sub-populations or classes is not directly observable to the researcher hence the term latent classes. Latent classes are considered to have a structure based on a set of observable variables, or covariates. The random relationships between covariates help determine the structure of latent classes. When confronted with a given set of choices – the attributes of a given housing stock – households will exhibit differences in residential location behavior due to differences in demographic and socioeconomic characteristics. Variations in demographic and socioeconomic characteristics and choice set response behavior occur across households and are considered unobservable to the researcher but influence household utility, purchase behavior and residential location decisions.

Unlike the random coefficient models, finite mixture models assume a discrete mixed distribution of the sample. I follow Cameron and Trivedi (2005) and assume a mixed distribution of:

$$y_i \sim \sum_{c=1}^{C} f(y_i|x_i, \beta_j)\pi_j \tag{4.1}$$

Where,

C= number of latent classes

β = coefficient vector for *N* explanatory variables

$\pi_j(z_j)$ = percentage chance of being associated with class C_i. Note that $\sum \pi_j = 1$

Note that, if a discrete mixed distribution is assumed then this implies that each observation can potentially have an associated probability function. Bearing this in mind, each observation from a sample can be regarded as a random draw from the *j*th

class. To represent this probability, d_{ij} is defined as the probability that individual *i* is in class j. That is,

$$d_i = (d_{i1}, d_{i2}, \ldots, d_{iC}) \text{ and } \sum d_{ij} = 1 \tag{4.2}$$

These unobserved probabilities are assumed to be independent and identically and distributed, with probabilities π, and multinomially distributed as,

$$(d_i|\pi) \sim \prod_{c=1}^{C} \pi_j^{\,d_{ij}} \tag{4.3}$$

The distribution of y_i given d_i becomes,

$$(y_i \mid d_i) = \prod_{c=1}^{C} \{ f(y_i \mid x_i, \beta_j)\pi_j \}^{d_{ij}} \tag{4.4}$$

Finite mixture methods reveal a probabilistic measure for each of the *N* total households into a latent class or type, given house values, structural attributes of each house, and the mixing covariates of each household. Class membership probabilities for each household and the beta coefficient for each class are simultaneously estimated given a predetermined estimate of C. These parameters are estimated by the likelihood function given by,

$$L = \prod_{i=1}^{N} h(y_i| x_i, \beta_j, \pi_j (z_i)) \tag{4.5}$$

Given all the above, the likelihood function is maximized by,

$$\text{Ln } L = \sum_{i=1}^{N} \sum_{c=1}^{C} [d_{ij} \ln f_j(y_i|x_i, \beta_j) + d_{ij} \ln (\pi_j)] \tag{4.6}$$

Note that *d* are treated as missing data since we are have not yet the existence of of submarkets in our area of study and we are unaware of the characteristics of those submarkets. If *d* was unknown, it would be estimated by maximizing the equation,

$$d_{ij} = \frac{e^{\,{}_j z_i}}{1 + \sum_{C=1}^{C} e^{\,{}_j z_i}} \tag{4.7}$$

Where,

z_i = vector of the C - 1 mixing covariates

$_i$ = gamma values; a weighting function on the demographic variables

However with unknown *d*, the estimation equation is then,

$$\hat{d}_{ij} = \frac{\pi_j f(y_i \mid x_i, \beta_j)}{\sum_{c=1}^{C} \pi_j f(y_i \mid x_i, \beta_j)} \tag{4.8}$$

Since the posterior probability of an observation belonging to a particular class is in not unknown, we use an EM algorithm to estimate these parameters. The Expectation step – the E step – involves imputation of the expected value of d_i given the mixing covariates, interim estimates of $_j$, β and π_j . The Maximization step – the M step – involves using estimates of d_i from the E step to update the component fractions of π_j and β. This process is repeated until there is no change in the likelihood function (4.6). Therefore, we simultaneously estimate class membership probabilities, gamma values of Z-vector, and beta coefficients. The EM step procedures can be summarized as follows:

1. Generate starting values for , β a nd π
2. Initiate iteration counter for the E-step, *t*(initial *t* at 0)
3. Use β^t and π^t from Step 2 to calculate provisional d^t and ^{t} with (4.7).
4. Initiate second iteration counter, *v*, for the M-step
5. Interim estimators of d^{t+1} are then used to impute new estimates of β^{v+1} and π^{v+1} with (4.8).
6. For each prescribed latent class, estimators of β^{v+1} are imputed, via the M-step, as well as π^{v+1}

7. Increase *v* counter by 1, and repeat M-step until:

$$f(\beta^{v+1}|y, x, \pi, d) - f(\beta^{v}|y, x, \pi, d) < \text{a prescribed constant; if so, then } \beta^{t+1} = \beta^{v+1} \quad (4.9)$$

8. Increase *t* counter and continue from step 3 until:

$$F(\beta^{t+1}, \pi^{t+1}, d|y) - F(\beta^{t}, \pi^{t}, d|y) < \text{a prescribed constant} \quad (4.10)$$

The steps above, particularly from Step 3-8 do not necessarily occur sequentially as outlined above but occur simultaneously as the continual updating of estimators. Each *v* iteration conditionally maximizes the likelihood function using interim estimates of observation latent class membership probabilities in one of the C latent classes; while each *t* iteration updates latent class memberships. Final estimates of β , π, and *d* when the convergence conditions (4.9) and (4.10) are met. These estimates are taken to be those most likely to account for observable outcomes.

The modified model to estimate implicit prices thus becomes:

$$\text{Price} = \beta_0 + \hat{d}_{ij} (\beta_N X_N) \quad (4.11)$$

Note that the number of classes is hypothesized prior to the implementation of the model. Notably, predominant research practice shows that the optimal number of classes can be set be a-priori information or selection can be based on information criterion such as Akaike information criterion (AIC), or the Bayesian Information Criterion (BIC) information criterions. I started by assuming 3 classes and compared the model fit, based on AIC, with 2 classes and 4 classes as well.

Moreover it is important to note that, hypothetically, if one were to increase the number of types from 4 to some large number *N*, the model would then mimic the theoretical implications of a random coefficients model. This implies that a finite mixture distribution would essentially collapse to a random coefficients distribution.

The random coefficient approach imposes continuity - it relies on the strong assumption that a type is distributed in a fairly continuous manner. A finite mixture approach does not impose this assumption.

Further note that with a finite mixture approach every individual will have a probability distribution of being in a certain type. The gamma's ( ) are weighting on the demographic variables such that every group will have a different weighting function. The effect of the demographic variables on the probability of being a certain type differs by type. This weighting function and the probability of being a certain type are done iteratively and simultaneously. Who is classified into what type then determines the beta coefficients for that type – these are jointly determined; and the beta coefficients also affect the probability of one being a certain type, this is done simultaneously. Upon convergence, every individual will have a probability distribution of being a certain type; finally note this probability distribution is could be thought of as being determined by a vector distance from the center of each type – the closer a household is to the center of a particular type the higher the probability you have of being that type. These probabilities are retained when determining the beta coefficients (or implicit valuations) for the different attributes. In this manner, a sharper estimate of attribute mix value is revealed.

CHAPTER V

RESULTS

5.1 The Data Sources

There are three sets of data used in this study, collected from different sources: housing residential sales data, survey data, percent tree cover and bird data. The first set of data collected was housing residential sales data from June 2006 to December 2008 within 17 different neighborhoods in Lubbock, TX; collected by extracting information from the Lubbock area Multiple Listing Service (MLS). From the MLS documents, I was able to record the final sales price of every house and the physical characteristics of every house. Additionally, I was able to extract the physical location of every sale from the MLS documents. With this information I was able to physically deliver surveys to every address that was listed as undergoing a sale transaction within that time period. I conducted three rounds of surveying and compiled the survey responses from each respondent into a distinct data set. The response rate for the survey was 51 percent, providing me 368 usable observations.

Overall, I collected a data set that was rich in house sales, tree canopy density and bird species richness within our neighborhoods of interest. Percent tree canopy cover was estimated from available GIS maps. Data on bird species abundance was collected by ecologists and wildlife scientists from Texas Tech University's Department of Natural Resources Management who conducted point count surveys where they observed and recorded birds at $n \geq 8$ sites in each of $n=17$ neighborhoods at ≥ 2 different mornings during summer 2009. They categorized bird species into three categories: Species *A* birds were common native and invasive species for example,

house sparrow (*Passer domesticus*), European starling (*Sturnus vulgaris*), great-tailed grackle (*Quiscalus mexicanus*), Eurasian collared dove (*Streptopelia decaocto*); Species *B* birds were desirable urban birds like: American robin (*Turdus migratorius*), blue jay (*Cyanocitta cristata*), mourning dove (*Zenaida macroura*), northern mockingbird (*Mimus polyglottos*), western kingbird (*Tyrannus verticalis*); while Species C birds were exurban or rare urban birds for example; meadowlarks (*Sturnella* spp.), pine siskin (*Carduelis pinus*), American goldfinch (*Carduelis tristis*), kinglets (*Regulus* spp.), and warblers (*Dendroica* spp. and others). By examining the behavior of these two variables in particular, we derived an effective instrument for measuring the effects of environmental externalities.

The bird variable used for analyses was calculated by multiplying the total number of birds by the total of bird species *B*. All data, such as tree canopy or the measure of the bird variable, were referenced to the residential unit of interest; so the bird variable, for example, realized a different value for each housing unit sale that we examined. Our approach is guided by a desire to examine these urban habitats to measure their ecological contributions, based on bird richness and relative abundance, and to construct a model to value these environmental externalities. We propose this instrumentation technique as a primer to valuate ecological outcomes of public greenspace and greenscapes. This instrumentation approach resembles other models that use birds as a barometer to assess urban ecosystem biodiversity or quality of greenspace such as Sandström et al (2006), Crooks et al. (2004) and Fernandez-Juricic (2001). Hence, there was theoretical grounding to use the bird variable as an

instrument; so we then sought to use the observed data on bird richness and relative abundance efficiently in incremental steps.

First, we sought to explain variation in the bird variable, defined as the total number of birds times the number of different kinds of Species B birds, through an efficient instrument. We regressed the bird variable against several variables including ‚near park,' a variable defined by the presence (1) or absence (0) of a park that met two criteria: the park was located within the neighborhood and also within a half mile of the residential unit of interest. Bird was also regressed against the percent tree cover and 16 dummy variables to capture variation due to neighborhood effects for the 17 neighborhoods defined. This regression was labeled (Instrumentation 1). We used the results from Instrumentation 1 and eliminated insignificant variables and then regressed the bird variable only against these significant variables to derive (Instrumentation 2). Following this, we used the results from Instrumentation 2 and eliminated those neighborhoods that had less than 6 sales; and we then regressed the remaining neighborhood dummy variables and tree cover against bird (Instrumentation 3). Following this, we compared the Instrumentations using AIC values for the different instrument regressions, comparing the AIC values for each model.
Having chosen the most efficient Instrument, we extracted the predicted value of bird diversity and abundance, Predicted Bird, for each home sale observed.

We collected MLS data for 368 home sales in Lubbock, TX from 2008-2009 and used 323 completed records for analyses. We conducted 296 point counts and used total number of birds and total number of species *B* detected on 2 (randomly selected if there were >2 conducted) point counts for each neighborhood to derive our

bird instrument. Instrument 3 Bird regressed against tree cover, significant neighborhood dummy variables with >6 sales was the best model. It showed the lowest overall AIC value; and all the explanatory weight is assigned to Instrument 3. We therefore used Instrument 3 to generate „predicted bird' (R^2_{adj}=0.4965) values for each house sale for which we had MLS data. Akaiki comparison of regression models used to develop instrument for deriving predicted bird variable. Variables used: bird, was derived using total birds times number of desirable bird species; near park, was distance to nearest open space/city park; and, neighborhood dummy variables. Data used in this regard are from summer 2009 in Lubbock, TX.

The demographic variables that were used in this analysis from the survey information were Age of the respondent, Gender of the respondent, the Household size of the respondent, the Education level of the respondent, and an attitudinal variable – the level of importance the respondents placed on the Crime Rate; together they formed the vector of covariates, *z*. The physical attributes of the house used in this analysis were Square Footage, Lot Size, and the age of the house; together with Predicted/Expected Bird, they formed the vector of attribute mix supply, *X*.

5.2 Econometric Results

The finite mixture approach tests the presence of differentiated households, and the effect of differentiation has on the attributes of a house and therefore house value as a whole. A perfect Tiebout model would dictate that people sort into neighborhoods based on preferences such that households with differing sets of preferences would not reside in the same neighborhood. By extension, a purely Tieboutian driven analysis would reveal neighborhoods with identical sets of

preferences; that the sorting into neighborhoods would driven by identical sets of preferences and not necessarily driven by demographic factors such as income. Residential planners approach residential design from the subscribe to viewpoint that demographic factors such as income drive residential choices such that households rank neighborhoods in the same order and as such use the same order of ranking in the residential choice decisions.

The implication of both approaches is that at households in the same neighborhood have identical preferences and they had identical order of rankings for neighborhoods and houses within the neighborhoods. Thus, when considering a house in a matrix of residential choices, whether they bought that house or not, households would consistently place that house in the same rank order within the decision space.

Conversely, we propose that when households valuate a house within an array of residential choices differently; such that households of different types could place a house in different ranks and, as a consequence, such rank neighborhoods differently. If this is true then, then their will exist inter-type differences in the implicit marginal prices for attributes of houses and neighborhoods. The differences will be a reflection of the differences in utility accrued from the varying neighborhood and house differences. Typing households in a statistically is a useful way of accounting for preference and utility differences in a structurally sound manner that offers deeper insight into the welfare of consumers of residential stock.

I accomplish the typing of households by implementing a finite mixture model which gives a probability distribution of an individual household being a particular type. In the case of a hedonic price model, one can think of the hedonic process as

assigning everyone to the same type with equal probability of valuing the housing attributes at the value of the beta coefficients produced by the model; especially since the coefficients of the housing variables form the best fit line for the whole data set. Thus, when, for example, at the small scale level of a landlord attempting to predict the change in value, or lack thereof, of increasing the square footage of a house; or at a larger scale of a planner weighing the change in value of building a residential areas with features that have positive ecology outcome, in both of these cases they would be a lose in efficiency and accuracy if the assignment of value to a feature of housing stock is constant across all households. Similarly, everyone would have the same predicted value for a particular house. This implies that if two individuals were competing for the same house, they would value each feature that makes up that particular bundle of housing stock; in such a case, the outcome of the bidding process parallel a Vickery process in that the individuals, seeking to acquire the house that maximizes utility, places bid until their maximum bid is reached – a process that is flooded with income driven undercurrents. As such, by extension the sorting process is purely income driven.

Again, this process is less efficient. A more efficient approach to setting the foundation of housing stock valuation would involve testing for, and if they so exist, identifying and separating utility functions for housing stock features across individuals. This will allow for a dependable improvement toward understanding how the valuation of features of housing stock differs across households; one of the important step to achieving this is the probability distribution per type per household. Every household has a probability of being a certain type; as one would expect, there

will be households that tend to strongly be a certain type while some households may weakly tend to be a certain type, while some may lie in within other ranges of probability. This is useful because when I calculate a predicted value for each household I am able to add to the value to the predictive information by proposing a probability of a household paying a certain price for house. The utility of this information goes beyond individual housing transactions and could be useful in exploring probable responses to a change in a feature of residential housing stock at a larger scope of reference. Furthermore, since the sorting criterions are the attitudinal and demographic characteristics of each household I am able to outline the general characteristics of a particular type.

Needless to say, this outcome is useful since typing households without being able to give some general characteristics of the household is less valuable.

The model best fitted the array of households into two types. I arrived at two types by hypothesizing that the population contained different numbers of latent classes, specifically 2, 3 and 4. Due to the size of the sample, I did not test for latent classes greater than 4. I recorded the AIC and Log-likelihood values for each iteration. The AIC value was computed during the mixing process and gave a measure of the overall goodness of fit of a given number of latent classes to the array of households. Therefore, the model with the lowest AIC value implies that that number of latent classes fits the array of households the best.

Additionally, in the surveying process, I collected demographic information. However, there is no criterion to use when choosing which demographic variables should or shouldn't be used. Therefore, I used common demographic variables as

sorting variables, i.e age, gender, household size education level which encompass a comprehensive description of a household; however, there is less direction on when or which attitudinal variables to include in a mixing process. Bearing this in mind, I run several rounds of the mixing model. Each round had a different combination of mixing variables. I recorded the AIC and log likelihood value for each round. Notably, I run each combination of mixing variables with a different number of latent classes (2-4) for consistency. In this way, I am able to not only deduce the best fit for the number of latent classes but also the mixing variables that best reveal and characterize the underlying latent classes in the array of households.

Notably, I implemented 16 different combination of mixing variables. The results from the different rounds were compared using the AIC basis. Out of the 16 iterations, I narrowed the combination of mixing variables to 6 combinations that had AIC and log-likelihood variables that were favorable in comparison to the other iterations. These 6 particular combinations had AIC and log likelihood values that were significantly different from the other combinations tested. This suggested that those 6 particular combinations were likely to be more stable and reliable in defining the latent structure than the other combinations. Hence, out of these 6 iterations, I choose the combination of mixing variables and number of types that had the best AIC value. Following this iterative process of comparison, I deduced a structure of best fit from the array of households rather than imposing a structure by assuming a certain number of types. Notably, out of all the different iterations of comparison across combinations of mixing variables and number of types, I deduced that the data set was

composed of two types with a latent structure described by their age, education level, gender, household size and their attitude toward crime.

The types exhibited differences in their attitudinal and demographic characteristics. The first type, which I will call Type 1, has a higher average income above the sample average income and the average income of the second type, which I will call Type 2: more than half of the Type 1 households reported earning an annual income of at least $125,000. Fifty-seven percent of Type 1 respondents were male, while 50% of Type 2 respondents were female. Type 1 households are well educated, with almost half of them achieving at least a graduate level education; this matches their attitudinal response of placing education higher than other potential responses (crime, environment, social security). Type 1 households had comparatively smaller households, cared more about environment than social security but tended to consider crime levels more than important either the environment and social security. On the other hand, Type 2 households have comparatively larger households, are less educated and are less wealthy. Almost 60% of Type 2 households earn between $50,000 to $125,000 annually while 13% of them earn less than $50,000 annually; additionally, almost two-thirds of Type 2 households have not achieved beyond an undergraduate level of education. Type 2 households care less about education than Type 1 households and are more inclined to care more about crime and social security than Type 2 households.

The survey responses indicated that more than one-third of Type 1 households indicated that „Tree Cover' of the neighborhood was one of the most important factor in decided to live in a neighborhood. More Type 2 respondents indicated that they

cared more about the investment potential and crime rate of the neighborhood than Type 1. This response seems reasonable as I would expect households with a larger size to be more sensitive to wealth generating opportunities and hence would their investments to have a higher long-term yield potential; additionally, with having a larger household size, presumably due to more children, one would expect them to be more wary of the crime rate and safety of their neighborhood. Further, more than half of Type 2 households indicated that the number of bedrooms was an important factor in influencing their purchasing decision of their particular house. Again, they are taking structural space into consideration, presumably because of a larger household size, more so than Type 1 households would on the margin.

Notably, similar to the guided iterative process of deducing an optimal combination of mixing variables and number of types, I eliminated insignificant housing sales price explanatory variables in the process of estimating a housing regression equation. As such I was left with four variables: square foot, lot size, house age, and (expected) bird. The implicit prices of these attributes of housing stock for each type were:

Table 5.2.1:

Implicit Valuation of Attribute-Mix by Type

	Type 1		**Type 2**	
	beta	se	beta	se
Constant	-28.6733	20.6999	6.928	6.0106
SqFt	0.0978	0.0067	0.0855	0.0022

Table 5.2.1. Continued.	**Type 1**		**Type 2**	
	beta	se	beta	se
Lot	1.6593	0.7146	1.6501	0.1718
HouseAge	-1.7506	0.4069	-2.3758	0.1302
ExpBird	0.747	0.2385	0.3159	0.0723

The results above generally follow my overall expectations, that is, I would expect a positive price effect for the structural variables of housing stock – square foot, lot size – and a negative effect of the age of the house. The results above imply that at the margin, Type1 households have a higher implicit value for square feet and the ecological features of a neighborhood; while Type 2 households have a lower implicit value for older houses and have similar marginal value for lot size. Type 2 households consider older houses as a strong disamenity and do not seem to value features in a neighborhood that have positive ecological outcomes; which implies that they prefer newer houses and open spaces with little landscape features. Generally, Type 1 households are more inclined to pay a premium to reside in neighborhoods that have ecological features that positive ecological outcomes; this propensity would resonate with their indication of tree cover as an important neighborhood amenity.

Using the above implicit values, I imputed a predicted value for each household. Needless to say, the probability distribution is embedded in the distribution of predicted values making the imputed predicted values a dynamic reflection of housing price from the given mixing distribution. The predicted value thus serve as a numeric composite representation of that household's utility measure of the house; a

utility that is a function of their preferences. Therefore, once we have the predicted values we can rank the predicted values across the whole data set for each type and compare how they rank each house; that is, we compare how well they perceive each house to match their preferences. Additionally, if we compare those predicted values at the neighborhood level, we can form inter-neighborhood comparisons to evaluate which neighborhoods are perceived to contain the highest value, the best utility, based on the median values of both types in that neighborhood. Table 2 below depicts the aforementioned comparisons.

Table 5.2.2:

Rankings of neighborhoods by type and percent composition of types per neighborhood

	Type 1	**%**	**Type 2**	**%**
1.	Rushland-Tanglewood	0.68	Rushland-Tanglewood	0.32
2.	Regal Park	0.09	Regal Park	0.91
3.	Lakeridge	0.5	Ravenwood	0.7
4.	Ravenwood	0.3	Lakeridge	0.5
5.	Southhaven	0.4	Southhaven	0.6
6.	Regency Park	0.12	Regency Park	0.88
7.	Melonies	0.28	Lakewood Estates	0.64
8.	Tech Terrace	0.82	Melonies	0.72
9.	Lakewood Estates	0.36	Pleasant Run	0.96
10.	Whisperwood	0.1	Whisperwood	0.9

Table 5.2.2. Continued.

	Type 1	**%**	**Type 2**	**%**
11.	Pleasant Run	0.04	Farrar	0.93
12.	Farrar	0.07	Tech Terrace	0.18
13.	Brentwood-Greenlawn	0.33	Brentwood-Greenlawn	0.67

Recall, the above coefficients shown in Table 1signify the implicit prices both types would pay for the given features of housing stock at the margin. Though it is a good starting point, the coefficients do not however give us a clear picture of how these implicit prices move when the two types come in direct competition such that they are forced to bid for the same house, or houses in the same neighborhood. To accomplish this, one would first have to identify the neighborhoods that both types have sizeable representation within them for a competitive process to be meaningful. Here, significant enough means a intra-neighborhood ratio of the two types that generally follows the ratio of types in the general array of households. The sample ratio of Type 1 to Type 2 is approximately 1:3. Bearing this in mind, using the composition of types in neighborhood as shown in Table 2 we can further identify and extract which neighborhoods these are. Table 3 shows these neighborhoods.

Table 5.2.3:

Neighborhood in which Types are in Competition

Type 1 (T1)	%	Rank(T1)	Type 2 (T2)	%	Rank(T2)
Rushland-Tanglewood	0.68	1	Rushland-Tanglewood	0.32	1

Table 5.2.3. Continued.

Type 1 (T1)	%	Rank(T1)	Type 2 (T2)	%	Rank(T2)
Lakeridge	0.5	3	Lakeridge	0.5	4
Melonies	0.28	7	Melonies	0.72	8
Ravenwood	0.3	4	Ravenwood	0.7	2
Lakewood Estates	0.36	9	Lakewood Estates	0.64	7
Brentwood-Greenlawn	0.33	13	Brentwood-Greenlawn	0.67	13

Table 5.2.4:

Median value of neighborhoods by Type

	Observed	T1, Predicted	T2, Predicted
Rushland –Tanglewood	360,000	416,895	345,612
Lakeridge	298,375	296,807	271,657
Melonies	164,500	225,415	177,615
Ravenwood	308,400	295,705	280,948
Lakewood Estates	158,495	188,569	189,921
Brentwood-Greenlawn	94,895	94,473	70,269

One of the ways to discern competition in these neighborhoods is to run a regression equation with a dummy variable that serves as a signal of whether or not an observation is in that neighborhood or not. However, I did not include Rushland-Tanglewood as this could distort the outcome since these neighborhoods had the highest median value in the sample and in the predicted values for both types as shown

in Table 5.2.4. The dummy variable would be included as an extra variable in the estimation of the modified hedonic price model. With this change, the output shown in table 5.2.5 was derived for the beta coefficients.

Table 5.2.5:

Implicit Valuation of Attribute-Mix by Type (when competing)

	Type 1		**Type 2**	
	beta	se	beta	se
Constant	-36.5195	21.18868	0.680812	5.356083
SqFt	0.098003	0.006546	0.079492	0.002135
Lot	1.403665	0.512557	3.70881	0.311402
HouseAge	-1.55856	0.421756	-2.51946	0.11566
ExpBird	0.600642	0.236008	0.314184	0.067476
Dummy	17.75834	13.72909	24.09541	3.025237

The results in Table 5.2.5 shed light on the effect of intra-neighborhood competition. Indeed, both types appear to change their bidding behavior when they are in these neighborhoods, or rather when they have to face competition from another type. Type 1househods do not seem to be bidding much differently for square footage, in other woods their marginal value for square feet is not a major driving factor in pushing them to attempt to outbid other households. However, the change in the beta coefficient for square footage for Type 2 households, from .079 to .085, imply that they are positively bidding on square footage; in other words, Type 2 households consider the square footage strongly in their decision to move to a particular house and

could potentially attempt to outbid other households solely based on the square footage of the house. Their willingness to seek to outbid other households based on square footage is in step with aforementioned indications that signal that they prefer larger structural size.

Additionally, even when in competition with other households, consider the age of a house as a disamenity and are likely to bid less for older houses. Type 1 households nonetheless are willing to pay a premium to live in a neighborhood that has features that have ecological value. This feature of neighborhoods is of high preference to Type 1 neighborhoods. With less urgency, they will positive bid on lot size to gain an edge in the bidding process and will also slightly bid less for newer houses. All in all, all the explanatory variables work in tandem with expectation except lot size for Type 2; additionally, the intercept term for Type 2 shifts considerably from .680812 to 6.928, both of these outcomes are hard to adequately rationalize.

A more refined way of deciphering the movements in marginal value in the face of household competition would be to implement a Monte Carlo-Markov Chain-Model Choice (MCMCMC), quantile-quantile analysis or a fixed effects model, and incorporate spatial effects in the mixture model. These would offer more reliable and revealing results that could be used to better understand valuation behavior during inter-type competition.

In this study, I hypothesized the potential existence of different types and then chose the appropriate number of types by comparing the AIC value. However, a more efficient approach would be to jointly determine the correct number of types, the beta

coefficients for those types and the probability distribution of each individual being a certain type. This approach, an MCMCMC model, would offer more accurate results. the fixed effects really is to improve the overall regression approach to sort neighborhood effects more cleanly and broadly; specifically, a spatial effects analysis with a weighting matrix by geo-coded position of fixed effects. In this regard, fixed effects and spatial econometrics would be broad improvements to sharpen estimates; the neighborhood dummy effect would be sharpened by fixed effects because it improves the regression as a whole. A Quantile-quantile regression could be used to show if houses where types have the same expected value for the unit pay a premium.

Overall, however, this study successfully integrated existing classical and neo-classical theory, in a well coordinated manner telling a coherent story that dovetailed into our empirical analyses. Therefore, the basic conclusions from this study can be summarized as follows: household residential sorting is possible using characteristics of households, and that sorting processes could lead to homogenous neighborhoods as well as heterogeneous neighborhoods. The study also exposes that residential models that rely on aggregation on attributes do not always bear reliable estimates. Additionally, I find that variables for mixing which were very consistent in isolating types are easily extracted, much less than the demands of the actual survey of the prior survey by Lipscomb and Farmer (2005); the approach is not as demanding for modest demographic and attitudinal information on households can be collected to sort households. This implies that the actual responses on an individual survey needed in order to achieve the computational demands of a finite mixture model are relatively small. Finally, while this study was only conducted for neighborhoods in Lubbock,

TX, this method shows promise for sparse selective surveys to sort and identify submarkets over a much larger region, say, the greater Los Angeles area et cetera. Therefore, the potential utility – from the viewpoint of terms of analytic extension and understanding of residential housing market structure – in applying a finite mixture approach is ample justification for the moderate extra effort of modest survey information collection in larger housing hedonic studies. The gains justify the means.

BIBLIOGRAPHY

Alfnes, F. (2007). Willingness to Pay versus Expected Consumption Value in Vickrey Auctions for New Experience Goods (2007). *American Journal of Agricultural Economics*, *89*(4), 921-931.

Abraham, J.M., Goetzmann, W.N. & Wachter, S.M. (1994). Homogeneous Groupings in Metropolitan Housing Market. *Journal of Housing Economics, 3*(3), 186–206.

Adair, A.S., Berry, J.N. & McGreal,W.S. (1996). HedonicModeling, Housing Submarkets and Residential Valuation. *Journal of Property Research, 13*(1), 67–83.

Bitter, C., Mulligan, G., & Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems*, *9*(1), 7-27.

Bourassa, S. C., Hoesli, M. & Peng, V. S. (2003) Do Housing Submarkets Really Matter?. *Journal of Housing Economics,* 12, 12-18.

Bourassa, S.C., Hamelink, F., Hoesli, M. & MacGregor, B.D. (1999). Defining Housing Submarkets. *Journal of Housing Economics, 8*(2), 160–83.

Boyle, M., & Kiel, K. (2001). A Survey of House Price Hedonic Studies of the Impact of Environmental Externalities. *Journal of Real Estate Literature*, *9*(2), 117-144.

Bruch, E., & Mare, R. (2009). Methodological Issues in the Analysis of Residential Preferences and Residential Mobility. (Unpublished Manuscript),University of Michigan, Ann Arbor, Michigan.

Cameron, A.C. & Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Can, A. (1990). The Measurement of Neighborhood Dynamics in Urban House Prices. *Economic Geography*. 66(3),254-271.

Chen, Z., Cho, S.-H., Poudyal, N., & Roberts, R. K. (2009). Forecasting Housing Prices under Different Market Segmentation Assumptions. *Urban Studies*. *46*(1), 167-87.

Crooks, K.R.,Suarez, A.V. & Bolger, D.T. (2004). Avian assemblages along a gradient of urbanization in a highly fragmented landscape. *Biological Conservation, 115*(3) 451–462.

Costello, G. (2001) A spatial approach to price segmentation in housing markets. Paper presented at the 8th European Real Estate Society Conference, Alicante, Spain, 27–29 June.

Dale-Johnson, D. (1982). An Alternative Approach to Housing Market Segmentation Using Hedonic Price Data. *Journal of Urban Economics, 11*(3), 311–32.

Dempster, A. P., Laird, N. M. and Rubin, D. B.(1977). Maximum Likelihood From Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society, 39B* (1), 1–38.

De-Palma, A., Picarda, N. and Waddell, P. (2007). Discrete Choice Models with Capacity Constraints: An Empirical Analysis of the Housing Market of the Greater Paris Region. *Journal of Urban Economics, 62*(2), 204–30.

Drew, D., & Skitmore, M. (2006). Testing Vickery's Revenue Equivalence Theory in Construction Auctions. *Journal of Construction Engineering & Management*, *132*(4), 425-428.

Ekeland, I., Heckman, J., & Nesheim, L. (2002). Identifying Hedonic Models. *American Economic Review*, *92*(2), 304-309.

Elidan, G., Lotner, N., Friedman, N., and Koller, D.(2001). Discovering hidden variables: A structure-based approach. *NIPS, 13*.

Epple, D. (1987). Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products, *Journal of Political Economy*, *95*(1), 59-80.

Farmer, M., Belasco, E., & Lipscomb, C. (2009). Using a Finite Mixture Model of Heterogeneous Households to Delineate Housing Submarkets. Presented at Southern Regional Science Association (SRSA). Alexandria, VA. March 18

Fernández-Juricic, E. (2000). Bird community composition patterns in urban parks of Madrid: the role of age, size, and isolation. *Ecological Research, 15*(4), 373–383.

Fernández-Juricic, E . (2001) Avian spatial segregation at edges and interiors of urban parks in Madrid, Spain. *Biodiversity and Conservation, 10*(8), 1303–1316.

Fik, T. J., Ling, D.C., & Mulligan, G.F. (2003). Modeling Spatial Variation in Housing Prices: A Variable Interaction Approach, *Real Estate Economics 31*(4). 623-646.

Friedman, N.(1998). The Bayesian structural EM algorithm. In Proceedings of the 14th Conference on Uncertainty in AI (129-138). San Francisco, CA: Morgan Kaufmann.

Goodman, A., & Thibodeau, T. (2003). Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics*, *12*(3), 181.

Goodman, A., & Thibodeau, T. (2007). The Spatial Proximity of Metropolitan Area Housing Submarkets. *Real Estate Economics, 35*(2), 209–32.

Grigsby, W. (1963). Housing Markets and Public Policy. Philadelphia, PA: University of Pennsylvania Press.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error, E*conometrica*, *47*, 153-61.

Hughes Jr., W., & Sirmans, C. (1992). Traffic externalities and single-family house prices. *Journal of Regional Science*, *32*(4), 487.

Kasahara, H., & Shimotsu, K. (2009). Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices. *Econometrica, 77*(1), 135-75.

Kauko, T. & Goetgeluk, R. (2005). Spatial and Multidimensional Analysis of the Dutch Housing Market Using the Kohonen Map and GIS. In ERSA Conference. Amsterdam, The Netherlands.

Kestens, Y., Thériault, M., & Des Rosiers, F. (2006). Heterogeneity in hedonic modelling of house prices: looking at buyers' household profiles. *Journal of Geographical Systems*, *8*(1), 61-96.

Kwak, S., Yoo, S., & Kwak, S. (2010). Valuing energy-saving measures in residential buildings: A choice experiment study. *Energy Policy, 38*(1), 673-677.

Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy*, *74*(2), 132.

Lazarsfeld, P.F. & Henry, N.W. (1966). Readings in Mathematical Social Science. Chicago: Science Research Associates Inc.

Lazarsfeld, P.F. & Henry, N.W. (1968). Latent Structure Analysis. Boston, MA: Houghton Mifflin Co.

Leishman, C. (2001). House Building and Product Differentiation: A Hedonic Price Approach. *Journal of Housing and the Built Environment, 16*(2), 131–52.

Lipscomb, C.A. & Farmer, M.C. (2005). Household Diversity and Market Segmentation within Single Neighborhood. *Annals of Regional Science, 39*(2), 791–810.

Luken, P.C. & Vaughan, S. (2006). Standardizing Childrearing through Housing. *Social Problems, 53*(3), 299-331.

Maclennan, D. & Tu, Y. (1996). Economic perspectives on the structure of local housing systems. *Housing Studies,11* 387–406.

Newsome, B. and J. Zietz. (1992). Adjusting Comparable Sales Using MRA - The Need for Segmentation. *Appraisal Journal*, 60, 129-35.

Nielsen, S.S., Grønbæk. C., Agger, J.F., & Houe, H. (2002). Maximum-likelihood estimation of sensitivity and specificity of ELISAs and faecal culture for diagnosis of paratuberculosis, *Prev. Vet. Med., 53*, 191–204

Nimon, W.R.,& Hall, R. (2007). Empirical Analysis of the Efficiency of Job Assignment Auctions. *Applied Economics, 39,* 1679-1690.

Palmquist, R.B. (2004). Property value models. In: Mäler KG, Vincent J (eds.) Handbook of Environmental Economics, 2, North-Holland.

Palmquist, R. (1984). Estimating the demand for the characteristics of housing. *Review of Economics & Statistics*, *66*(3), 394.

Quandt, R. (1956). A probabilistic theory of consumer behavior. *Quarterly Journal of Economics, 70,* 507-536.

Quandt , R. (1972). A New Approach to Estimating Switching Regressions. *Journal of the American Statistical Association, 67* (338), 306-10.

Quigley, J.M. (1985). Consumer Choice of Dwelling, Neighborhood and Public Services. *Regional Science and Urban Economics, 15*(1), 41–63.

Riddel, M. (2001). A Dynamic Approach to Estimating Hedonic Prices for Environmental Goods: An Application to Open Space Purchase. *Land Economics*, *77*(4), 494.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82, 34-55.

Rothenberg, J., Galster, G., Butler, R. & Pitkin, J. (1991). *The Maze of Urban Housing Markets: Theory, Evidence and Policy* (Chicago, University of Chicago Press).

Samuelson, P.A. (1954). The Pure theory of public expenditure. *The Review of Economics and Statistics, 36 (4)*, 387-389.

Sandström, U. G., Angelstam, P., & Mikusiński, G.(2006). Ecological diversity of birds in relation to the structure of urban green space. *Landscape & Urban Planning, 77*(1/2), 39-53.

Schnare,A.B., & Struyk, J. R. (1976, April). Segmentation in urban housing markets *Journal of Urban Economics*, *3*(2), 146-166.

Sirmans, G., MacDonald, L., Macpherson, D., & Zietz, E. (2006). The Value of Housing Characteristics: A Meta Analysis. *Journal of Real Estate Finance & Economics*, *33*(3), 215-240.

Straszheim, M. (1975). An Econometric Analysis of the Urban Housing Market. National Bureau of Economic Research. New York, NY: Columbia University Press.

Stigler, G. J. (1945). The Cost of Subsistence. *Journal of Farm Economics*, 27.

Suzmumura, K. & Xu, Y. (2003). Recoverability of choice functions and binary relations:some duality results. *Social Choice & Welfare, 21*(1).

Teratanavat, R., & Hooker, N. (2006). Consumer Valuations and Preference Heterogeneity for a Novel Functional Food [electronic resource]. *Journal of food science, 71*(7), 533-541.

Thatcher, J., Morey, E., and Craighead, W.E. (2005). Using patient characteristics and attitudinal data to identify depression treatment preference groups: a latent-class model. *Depression and Anxiety, 21*(2), 47-54.

Thünen von, J.H.(1966). *von Thünen's Isolated State*.(P. Hall, Trans.). Oxford, U.K.: Pergammon. (Original work published 1826).

Tideman.Published in Russian in Alexander Pozdnyakov (ed.) Order and Chaos in the Development of Socio-Economic Systems, Tomsk: The Institute of Optical Monitoring, 1998, pp. 16-25. The Ethics of Coercion in Public Finance. Nicolaus Tideman. Professor of Economics. Virginia Polytechnic Institute and State University

Tiebout,C. (1956). A pure theory of local expenditures. *Journal of Political Economy, 64 (5)*, 416–424.

Tu, Y. (1997). The Local Housing Submarket Structure and Its Properties. *Urban Studies, 34*(2), 337–53.

Tu, Y., & Goldfinch, J. (1996). A Two-stage Housing Choice Forecasting Model. *Urban Studies, 33(3)*, 517-538.

Vickery, W.J. (1961). Counterspeculation, Auctions, and Competitive Sealed Tenders. *Finance*, *16.*

Wallmo, K., & Edwards, S. (2008). Estimating Non-market Values of Marine Protected Areas: A Latent Class Modeling Approach. *Marine Resource Economics, 23*(3),301–323.

Watkins, C.A. (2001). The Definition and Identification of Housing Submarket. *Environment and Planning ,A 33*(12), 2235–53.

Zhuo, C., Seong-Hoon, C., Neelam, P., & Roberts, R.K.(2009). Forecasting Housing Prices under Different Market Segmentation Assumptions. *Urban Studies, 46*(1), 167-187.